

周报总结

2015-07-06

本周工作：

日程安排：

周一：

上午：查找与心理健康有关的论文与资料，完成论文的 800 字。

下午：上课。

周二：

上午：完成《心理健康课程》大论文，800 字-3000 字部分

下午：完成《心理健康课程》大论文，3000 字-5000 字部分

周三：

上午：查找自然辩证法课程相关考试资料。

下午：自然辩证法考试

周四：

上午：完成《心理健康课程》大论文，5000 字-6000 字部分，整理格式

下午：《心理健康课程》期末随堂测验，上交大论文

周五：

修改系统，跟巫老师讨论。

具体工作

跟巫老师讨论相关系统之前的工作后，老师提出意见，聚类可以不用全部属性聚类，而是只对部分感兴趣的属性进行聚类（感兴趣的属性可以设置成比较相关的一些类，比如就玩家实力而言，我们可以选取玩家的：装备等级，修炼，修为等切实反应实力的属性）。这样聚类结果可能会更加反应一些问题。

讨论后，因为数据较大，因此可能聚类时间可能较长，所以在此先选取部分的时间的数据进行聚类，看一下效果十分令人满意。

选取的时间为 2013 年 5 月 24 日到 2013 年 5 月 31 日将近一周的数据。属性选取了三个组

每组大概 3-5 个属性，因为我们选取的属性并不多，所以 K 值取值没有选则的太大，这里选取的 k 值为 3、5、8、10 四个 K。大致的聚类结果如下

```
abilityAttr_20130524_20130531_K3.txt
abilityAttr_20130524_20130531_K3_center.txt
abilityAttr_20130524_20130531_K5.txt
abilityAttr_20130524_20130531_K5_center.txt
abilityAttr_20130524_20130531_K8.txt
abilityAttr_20130524_20130531_K8_center.txt
abilityAttr_20130524_20130531_K10.txt
abilityAttr_20130524_20130531_K10_center.txt
moneyAttr_20130524_20130531_K3.txt
moneyAttr_20130524_20130531_K3_center.txt
moneyAttr_20130524_20130531_K5.txt
moneyAttr_20130524_20130531_K5_center.txt
moneyAttr_20130524_20130531_K8.txt
moneyAttr_20130524_20130531_K8_center.txt
moneyAttr_20130524_20130531_K10.txt
moneyAttr_20130524_20130531_K10_center.txt
timeAttr_20130524_20130531_K3.txt
timeAttr_20130524_20130531_K3_center.txt
timeAttr_20130524_20130531_K5.txt
timeAttr_20130524_20130531_K5_center.txt
timeAttr_20130524_20130531_K8.txt
timeAttr_20130524_20130531_K8_center.txt
timeAttr_20130524_20130531_K10.txt
```

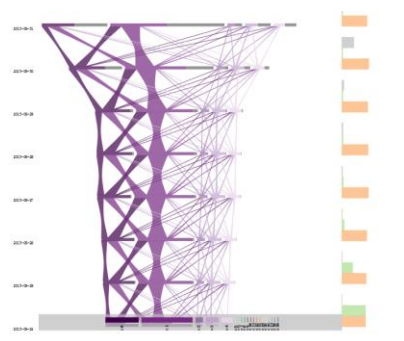
这其中，timeAttr 描述了玩家跟时间有关的属性，比如上上限次数，当天上线时间，玩家年龄等。

Money 属性主要包含了玩家充值金额，玩家银两数量，玩家元宝数量，玩家银票数量等跟金钱有关的属性。

Ability 属性主要包含了跟玩家能力有关的属性，比如：装备等级、修炼等级、修为等级等。进行聚类之后，为了能够展示，还要修改系统，因为过去的系统并不能兼容只有部分数据被聚类之后的情况。

以下展示一下几种不同聚类之后的结果：

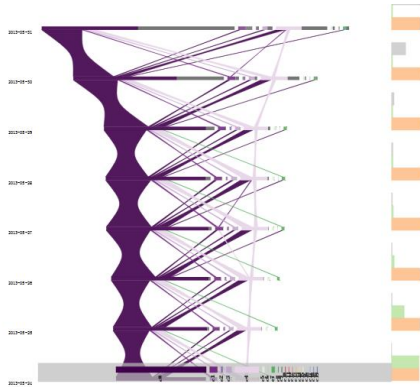
玩家能力（5 类）





可以看到，聚类少了之后，不同类别之间的转移明显增多，有的类别之间交叉现象十分明显。

跟金钱有关的类别转移：



玩家时间相关属性（8类）

在做这部分属性的时候，想到一个问题当时想做没做的交互，由于自身类别转移较多，而我们往往更加关注的是向他人的转移，因此，我们可以加入一些交互，从而屏蔽自身的转移，这样能让我们的系统更加清晰。

尚存问题：

由于之前系统并没有针对部分聚类结果进行展示，因此，要兼容部分展示，还要对系统进行一定程度上的修改。比如现在只能看转移，还不能看具体其中的类别数据分布，这是因为，我们之前进行展示分布的时候，为了提高展示的效率，我们数据采用了预处理+实时计算的方法，部分数据预处理，部分数据实时计算，这也导致，要想兼容部分属性的展示，就必须修改系统（系统程序过去对数据的耦合很强）。希望下周能解决这个问题。